Excercise 1 – Hypothesis classes and XOR

Yoav Goldberg

Consider the linear hypothesis class:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$$

Our goal is to represent the XOR function, defined as:

$$xor(0,0) = 0$$

 $xor(1,0) = 1$
 $xor(0,1) = 1$
 $xor(1,1) = 0$

Here, \mathbf{x} is a 2-dimensional vector where each dimension takes the values of 0 or 1 only.

You can assume any decision rule you want, but for simplicity lets take the decision rule to be sign(f(x)), that is, the answer is 1 if $f(x) \ge 0$ and 0 if f(x) < 0.

Q1

Can you find parameters (that is, a 2-dim vector \mathbf{w} and scalar b) such that $\mathbf{w} \cdot \mathbf{x} + b = \operatorname{xor}(\mathbf{x})$? If not, why not?

Q2

Now consider the 2-layers hypothesis class:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{h} + b_1$$
$$\mathbf{h} = \max(\mathbf{U}\mathbf{x} + \mathbf{b}_2, 0)$$

where **w** is a 2-dim vector, b_1 a scalar, **U** is a 2×2 matrix, **b**₂ is a 2-dim vector, and **h** is a 2-dim vector. The max function is applied element-wise to the 2-dim result of **Ux** + **b**.

a. Find parameters (values for $\mathbf{w}, \mathbf{U}, \mathbf{b}_2, b_1$) such that $sign(f(\mathbf{x})) = xor(\mathbf{x})$.

b. Can you do this without the max function? Can you show it is not possible?